

Requested Patent: JP2003018188A

Title:

FLOW ARCHITECTURE FOR REMOTE HIGH-SPEED INTERFACE APPLICATION ;

Abstracted Patent: JP2003018188 ;

Publication Date: 2003-01-17 ;

Inventor(s):

AYDEMIR METIN;; RINDOS III ANDREW J;; RODRIGUEZ JORGE ROLANDO ;

Applicant(s): INTERNATL BUSINESS MACH CORP ;

Application Number: JP20010195009 20010627 ;

Priority Number(s): ;

IPC Classification: H04L12/56; G06F13/36 ;

Equivalents: ;

ABSTRACT:

**PROBLEM TO BE SOLVED:** To solve the problem of there being the possibility that a serial queue, such as a PCI bus interface, is turned to be a bottleneck resulting in the latently annihilating benefits on performance for use of a high-speed network switching interface. **SOLUTION:** This system having an inter-remote bus high-speed switching interface is provided with a switching mechanism, to which a plurality of bus interfaces are connected. A programmable flow queue, having a plurality of parallel logical flow queues, is used for scheduling a packet in accordance with the protocol request of a remote bus interface.

(19)日本国特許庁 (JP)

## (12) 公開特許公報 (A)

(11)特許出願公開番号

特開2003-18188

(P2003-18188A)

(43)公開日 平成15年1月17日(2003.1.17)

(51)Int.Cl.	識別記号	FI	キーワード(参考)
H04L 12/56		H04L 12/56	F 5B061
G06F 13/36	310	G06F 13/36	310C 5K030

審査請求 有 請求項の数13 OL (全 8 頁)

(21)出願番号 特願2001-195009(P2001-195009)

(22)出願日 平成13年6月27日(2001.6.27)

(71)出願人 390009531

インターナショナル・ビジネス・マシー  
ズ・コーポレーションINTERNATIONAL BUSIN  
ESS MACHINES CORPO  
RATION

アメリカ合衆国10504、ニューヨーク州

アーモンク ニュー オーチャード ロー  
ド

(74)代理人 100086243

弁理士 坂口 博 (外1名)

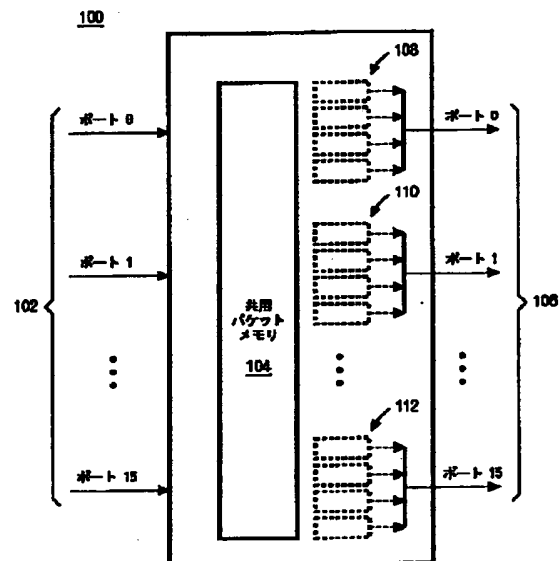
最終頁に続く

(54)【発明の名称】 リモート高速インターフェースアプリケーション用フローアーキテクチャ

## (57)【要約】

【課題】 PCIバスインターフェースのようなシリアルキューは、高速ネットワークスイッチングインターフェースを用いることのパフォーマンス上の利点を潜在的に減殺しうようなボトルネックになる可能性がある。

【解決手段】 リモートバス間的高速スイッチングインターフェースを有するシステムであって、多数のバスインターフェースが接続されるスイッチ機構を含む。リモートバスインターフェースのプロトコル要求に従ってパケットをスケジューリングするために、多数の並列論理フローキューを有するプログラム可能なフローキューが用いられる。



## 【特許請求の範囲】

【請求項1】 リモートバス間の高速スイッチングインターフェースを提供するための装置において、スイッチ機構と、リモートバスと、前記リモートバスを前記スイッチ機構にインターフェースする、リモートバスインターフェースと、前記リモートバスインターフェースのプロトコル要件に応じて、パケットのスケジューリングを行うためのプログラム可能フローキューを含む、装置。

【請求項2】 前記プログラム可能フローキューは、前記スイッチ機構の中に組み込まれており、前記リモートバスインターフェースから前記スイッチ機構に到着したパケットが、前記スイッチ機構から出る前に、前記プログラム可能フローキューへ経路指定される、請求項1記載の装置。

【請求項3】 前記リモートバスインターフェースが、リモート周辺コンポーネントを相互接続するインターフェースである、請求項1記載の装置。

【請求項4】 前記リモートバスは、前記リモートバス間でデータの順序を決定するためのバスアーキテクチャを有し、かつ、前記プログラム可能フローキューは、更に、前記バスアーキテクチャに従ってパケットのスケジューリングを行うための、データ順序付けロジックを有する、請求項3に記載の装置。

【請求項5】 前記プログラム可能フローキューは、複数のパラレル論理フローキューを含む、請求項1に記載の装置。

【請求項6】 前記データ順序付けロジックは、前記バスアーキテクチャに従い、前記複数のパラレル論理フローキューを特徴付ける、請求項5に記載の装置。

【請求項7】 前記プログラム可能フローキューは、更に、複数のパラレル論理フローキューのそれぞれに対するキューしきい値要件を、個別にプログラマブルに調整するためのロジックを有する、請求項6に記載の装置。

【請求項8】 前記スイッチ機構は、前記リモートバスからの入力パケットを受信するためのソースポートを有し、かつ、前記装置は更に、前記バスアーキテクチャに従い、前記入力パケットを前記プログラム可能フローキューに経路指定するためのバスアーキテクチャ命令を保持するトランザクション作業キューを含む、請求項4に記載の装置。

【請求項9】 スwitch機構を介してリモートバス間の高速スイッチングインターフェースを提供するための方法であって、前記方法は、前記スイッチ機構によってリモートバスをインターフェースし、前記リモートバスインターフェースのプロトコル要求に従いパケットをスケジューリングする、方法。

【請求項10】 前記スイッチ機構にはプログラム可能

フローキューが含まれており、入力パケットが前記スイッチ機構を出る前に前記入力パケットを前記プログラム可能フローキューに経路指定するステップを更に含む、請求項9に記載の方法。

【請求項11】 前記リモートバスは、前記リモートバス全体にわたりデータ順序を決定するためのバスアーキテクチャを有し、前記方法は、前記バスアーキテクチャに従いパケットをスケジューリングするステップを更に含む、請求項10に記載の方法。

【請求項12】 前記プログラム可能フローキューは、複数の並行論理フローキューを含み、前記複数の並行論理フローキューのそれぞれに対して独立にキュー閾値要求を調節するステップを更に含む、請求項10に記載の方法。

【請求項13】 前記スイッチ機構は、前記リモートバスからの入力パケットを受信するためのソースポートを有し、前記バスアーキテクチャに従って前記入力パケットを前記プログラム可能フローキューに経路指定する、請求項10に記載の方法。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】本発明は、一般的には通信ネットワークを管理するための方法及びシステムに関し、特に通信ネットワーク内のスイッチングフロー制御機構に関する。更に詳細には、本発明は、従来の優先度付け出力キューイングを、論理フロー制御機構に置き換えるための方法及びシステムに関する。この論理フロー制御機構は、データ順序付け及び関連するフロー制御が同時に処理され得るように、リモートバス相互接続として用いることの可能なスイッチ機構の内部に実装される。

## 【0002】

【従来の技術】従来、遅いネットワーク速度、及び伝送制御プロトコル/インターネット・プロトコル(TCP/IP)のようような上位層プロトコルにより提供されるフロー制御が、スイッチにおけるフロー制御上の問題を回避してきた。近年のノード間フロー制御や通信ネットワーク内のデータ伝送媒体の発達により、ブリッジ若しくはスイッチにおける輻輳がより大きな問題になってきた。全二重伝送機能に呼応した近年の1ギガビット/秒に及ぶネットワーク速度増加に応じて、フロー制御機構が要求されてきている。このようなフロー制御の目標には、効率とフェアネスが含まれる。

【0003】サーバー入出力(I/O)は、次世代I/O(NGIO)、フューチャーI/O(FIO)、及びシステムI/O(SIO)のようようなニューイニシアティブにより定義された方向性で発展している。NGIO、FIO及びSIOはInfiniBandコンソーシアムによってInfiniBandアーキテクチャへと統合された。これらのI/Oアダプターは、ホストチャネルアダプター(HCA)とターゲットチャネルアダプタ

ー(TCA)を相互接続するための、スイッチ機構を要する。

【0004】インターナショナル・ビジネス・マシーンズ(IBM)製のPRIZMAスイッチ構造のような、多くの従来型スイッチ機構は、プログラム可能なバケットフロー優先度方式を用いている。初期化の際に、利用可能な優先度の数として1からnの値を設定できる。したがって、このようなスイッチは、n個のキューによる優先順位付け方式による最大nの優先度レベルをサポートすることが可能である。スイッチング機構の入出力部において、異なる優先度を持つ複数のバケットのフロー制御を行うために、フロー制御機構を用いることができる。このようなスイッチ機構において入力フロー制御を実現するのに、オンチップピンを用いても良いし、或いは各バケットのヘッダー内の受信許可(receive grant)情報によっても良い。スイッチ機構の出力におけるフロー制御は送信許可ピンにより実現可能である。

【0005】このような優先度付きフロー制御方式においては、ある優先度が使用不可にされるとそれより下位の優先度の全てもまた使用不可とされることになる。n個の優先度レベルに対応したn個のキューは次のような優先度規則によりサービスされる。すなわち、高い優先度のキュー内で待機中のバケットは、どの下位の優先度キューよりも前に転送される。

【0006】

【発明が解決しようとする課題】従来型スイッチ機構内におけるフロー制御しきい値及びバックプレッシャートリガーは、優先度ベースのキューイングに依拠しており、個々のキューサイズに対して累積出力キューサイズを用いていた。ゆえに、高い優先度が使用不可にされた時にはいつも、それより下位の全ての優先度のトラフィックは使用不可になることが保証されていた。PCIバスインターフェースのようなシリアルキューは、高速ネットワークスイッチングインターフェースを用いることのパフォーマンス上の利点を潜在的に減殺しうるようなボトルネックになる可能性がある。PCIは、よく知られた32/64ビットローカルバス標準であり、パーソナルコンピュータ内の、ユーザーディスプレイ、ディスクドライブ、モデム等の周辺機器に高速にアクセスするために設計されたものである。

【0007】スイッチI/Oにおけるバスインターフェースの物理的割り込みによって、スイッチ内部のバッファオーバーフロー若しくはアンダーフローを防止するために適切なバッファ管理が施されなければならない。リモートPCIバスインターフェースのようなシリアルキューはフロー制御の潜在的なボトルネックになる。したがって、スイッチのキューイング能力を利用して、リモートバスインターフェースが遠隔切り替えできるような、パラレルなフロー制御とデータ順序付け機構を維持

するシステム及び方法に対する要求が存在することが理解されよう。

【0008】

【課題を解決するための手段】以下、リモートバス間の高速スイッチングインターフェースを有するシステムを開示する。本システムは、複数のリモートバスインターフェースが接続されるスイッチ機構を含む。リモートバスインターフェースのプロトコル要件に応じてバケットのスケジューリングを行うために、複数のパラレル論理フローキューを含む、プログラム可能フローキューが用いられる。すなわち、本発明によれば、リモートバス間の高速スイッチングインターフェースを提供するための装置において、スイッチ機構と、リモートバスと、当該リモートバスを前記スイッチ機構にインターフェースする、リモートバスインターフェースと、リモートバスインターフェースのプロトコル要件に応じて、バケットのスケジューリングを行うためのプログラム可能フローキューとを含む、装置が提供される。プログラム可能フローキューは、スイッチ機構の中に組み込まれており、リモートバスインターフェースからスイッチ機構に到着したバケットが、スイッチ機構から出る前に、プログラム可能フローキューへ経路指定されても良い。また、リモートバスインターフェースが、リモート周辺コンポーネントを相互接続するインターフェースであっても良い。更に、リモートバスは、リモートバス間でデータの順序を決定するためのバスアーキテクチャを有し、かつ、プログラム可能フローキューは、更に、バスアーキテクチャに従ってバケットのスケジューリングを行うための、データ順序付けロジックを有しても良い。ここで、プログラム可能フローキューは、複数のパラレル論理フローキューを含んでも良い。データ順序付けロジックは、バスアーキテクチャに従い、複数のパラレル論理フローキューを特徴付けても良い。プログラム可能フローキューは、更に、複数のパラレル論理フローキューのそれぞれに対するキューしきい値要件を、個別にプログラマブルに調整するためのロジックを有しても良い。スイッチ機構は、リモートバスからの入力バケットを受信するためのソースポートを有し、かつ、当該装置は更に、上記バスアーキテクチャに従い、入力バケットをプログラム可能フローキューに経路指定するためのバスアーキテクチャ命令を保持するトランザクション作業キューを含んでも良い。また、本発明は、スイッチ機構を介してリモートバス間の高速スイッチングインターフェースを提供するための方法であって、前記スイッチ機構によってリモートバスをインターフェースし、前記リモートバスインターフェースのプロトコル要求に従いバケットをスケジューリングする方法も提供する。本発明の目的、特徴及び利点については以下に示す詳細な説明で明らかとなろう。尚、本発明の特徴と思われる新規な点については特許請求の範囲に掲げるが、発明自体に加

え、好ましい使用形態、別の目的及びその利点に関しては、以下の実施例の詳細な説明を図面と併せて参照することで、最も理解されるであろう。

#### 【0009】

【発明の実施の形態】本発明は、従来のスイッチ内フローの優先度方式を論理制御アーキテクチャに置き換えるものであり、従来の $n$ 個の優先度キューが、 $n$ 個の論理的フローキューに置き換えられる。PCI間スイッチング等の実装をサポートするための、1ポートに対して多数のフローを有するスイッチング機構の構成を開示する。本発明は、新たに開発されたPRIZMAスイッチ（IBM製）内に見られるような従来型スイッチ機構に改変を加え、数多くの様々なトラフィックフローを効率的にサポートできるようにする。

【0010】このような新しい出力キューアーキテクチャによって、複数の論理フローに対する個別のサービス及びフロー制御が可能となる。このような柔軟性のあるキューイングアーキテクチャは、例えば、PCIバス順序付け命令が遵守されなければならないような実装を行う場合に要求される。各論理フローは、従来の優先度キューと同様に物理出力ポートにキュー（即ち、論理ポート）を有する。これら論理キューは伝送に関し、個別にかつ独立に使用可、使用不可にすることができる。従来の優先度キューと異なり、これらの論理フローキューは固有の相互に独立な順序付け方式によってプログラムされてはいない。物理出力ポートは、要求される処理に従いプログラマブルに決定された順序で論理フローキューの出力を提供する。例えば、論理フローキューの集合体から物理出力ポートへの出力フローは、ラウンドロビン方式で行われてもよく、またPCIバスアプリケーションの場合には、フロースケジューリングはPCIバスコマンド順序付けの規則により確立された要件を実現するようにプログラムされるであろう。

#### 【0011】

【実施例】以下図面を参照するに際し、全図面を通し、同じ物及び対応するパートに対しては同じ参照番号を付す。図1において、通信スイッチ100（スイッチ機構“switch fabric”と呼ぶ場合もある）内に本発明によるフローアーキテクチャが実装されている。図示された実施例において、スイッチ100は、16の双方向ポートを有するバスベースの16×16スイッチングユニットである。ポートの双方向性を簡明に表すために、16の出力ポート106に対応する16の入力ポート102を示す。スイッチ100はPRIZMAスイッチであっても良く、各ポートは各方向に2Gbps若しくはそれより大きいバンド幅を提供することが可能である。このようなポート速度にすると、スイッチ100が、最近の高速ネットワークI/Oで要求されるバンド幅を提供するに特に適したものとなる。

【0012】更に図1に示される通り、スイッチ100

は共用バスメモリモジュール104を含む。メモリモジュール104には、あらゆるタイプの様々なコンピュータメモリ装置を用いて良い。図2及び3を参照して更に詳細に説明するように、複数の入力ポート102の1つから入力される入力バスケットはプログラム可能である。これに関しては、図示された実施例においては、バスケットサイズは32バイトであり、当該バスケットは、3バイトのヘッダーとデータフィールドを含むとする。宛先ポート（複数の出力ポート106のうちの1つ）が、バスケットヘッダー中に定義される。宛先ポートアドレスをビットマップフォーマットで用意することにより、何れの出力ポート106からのマルチキャストリングも可能である。スイッチ100は、一組の論理フローキュー内に実現されたプログラム可能なバスケット送信（キューイング）を有する。当該一組の論理フローキューとは、出力ポート0に関連付けられた論理フローキュー108、出力ポート1に関連付けられた論理フローキュー110及び出力ポート15に関連付けられた論理フローキュー112などを含む。

#### 【0013】メッセージの構造

図2に、本発明の好適な実施例による、スイッチバスケットの内部構造を示す。以下で詳細に説明を行うが、スイッチ100はリモートバスからのサービス若しくはアクセス要求にตอบสนองしてトランザクションバスケット200を構成する。図2に示される通り、トランザクションバスケット200は、トランザクションヘッダー208を含む。トランザクションヘッダー208は、スイッチ100を介して正しい出力（ターゲット）ポートにトランザクションバスケットを導くためのスイッチ内部経路に関する経路情報を含む。トランザクションバスケット200は更に、トランザクションペイロード206を含む。トランザクションペイロード206は、ターゲットバスインターフェースロジック（不図示）により解釈され実行されるべきメッセージを含む。PCI間スイッチングにおいては、このようなバスインターフェースロジックは、ターゲットバスPCIシーケンサーということになる。

【0014】トランザクションバスケット200内のオリジナルメッセージは、メッセージヘッダー204及びメッセージペイロード202を含む。メッセージヘッダー204は、トランザクション固有の情報（例えば、コマンド、アドレス、バイトイネーブル）を含み、メッセージペイロード202は最大32バイトのデータを含む。32バイトよりも少ないペイロードデータを持つメッセージが、トランザクションヘッダー208内部にタグされても良い。このタグは、上記したターゲットバスインターフェースロジック（シーケンサー）により利用され、単一若しくはそれ以上のデータワードを転送するための一連のメモリ書き込みコマンドが生成される。

#### 【0015】キュー構造

図3には、論理ポートアダプター300が示されている。当該アダプターは、入力ポートから物理出力ポートへ転送されるバケットを本発明の好適な実施例に従って、キューイングするためのものである。本発明のキューイングシステム及びキューイング方法のキーとなる特徴の一つは、フレキシブルな経路指定及び出力キューイングがなされることで、このことはどのような特定のリモートバスアーキテクチャにも適用可能である。図示及び説明の簡便のために、以降の図面はPCI間スイッチング構成であるものとして扱う。

【0016】論理ポートアダプター300は、出力ポートキューイングを行う、スイッチ100のスイッチ構造（スイッチングコアと呼ぶこともある）内に実装されている。したがって、全ての論理フローキュー108乃至112は、スイッチ100の“下流”に向けた出力を有する。図3aはスイッチコアに実装された論理フローキューの構成を示す。詳細には、論理フローキューQ1、Q2、Q3及びQ4は、プログラム可能フローキューブロック302を構成している。このフローキューブロック302には出力ポート310が関連付けられ、更に出力ポート310には、リモートバスBが関連付けられている。同様に、論理フローキューQ1'、Q2'、Q3'及びQ4'はプログラム可能フローキューブロック304を構成し、キューブロック304は、リモートバスAの入力ポート308と関連付けられている。図示された実施例では、バスA及びBは双方向的であって良く、したがって、同時にソースとしてもデスティネーションとしても働き得る。図3には二つのバスしか描かれていないが、論理ポートアダプター300は、単一若しくはそれ以上のソースバス及び単一若しくはそれ以上のデスティネーションバスを有する構成にも適用し得る。

【0017】トランザクションバケット200のような、トランザクションバケットであって、バスBに転送されるべきものは、プログラム可能論理フローキューブロック302を介して出力ポート310に送信される。各々のタイプのトランザクションに割り当てられるキューは以下の基準に従って決定される。第1に、システムを通過するトランザクションは、バスA及びバスBの一方若しくは双方により使用されているデータ順序付け規則を満たすこと。次に、数多くのトランザクションを独立した並行な複数の経路中を移動させると、より良いフロー及びバッファリング効率が得られるということである。図4は、キューの割り当てに関し、上記第1の基準に沿い、図3の論理アダプター内に実装し得るバスアーキテクチャ仕様規則を図示したものである。この例では、論理ポートアダプター300を通過するバケットデータは、図4に記載のデータ順序付け規則を満たす。図中、行→列は、許容される転送を示す。本図に示される仕様規則は、以下図5及び図6に示される実施例におけるキューイング方式及びコマンド操作の説明に用い

る。

#### 【0018】コマンド操作

図5は、本発明によるスイッチアーキテクチャ400を示す。スイッチアーキテクチャ400は双方向PCIバスA及びBと、これらに対応するPCIインターフェース404及び418を含む。スイッチアーキテクチャ400は更に、バスBの出力ポートに関連付けられたプログラム可能フローキュー412及び、バスAの出力ポートに関連付けられたプログラム可能フローキュー410を含む。プログラム可能フローキュー412及び410は、複数の論理フローキューを含む点で、プログラム可能フローキュー302及び304と類似する。更に、図5に示される通り、スイッチアーキテクチャ400は、トランザクション作業キュー（TWQ）406及びTWQ416を含む。これらTWQは、プログラム可能フローキュー410及び412を介してPCIインターフェース418及び404間に転送されるべき入出力バケットを処理する。以下には、双方向コマンドフロー及びデータにおいて使用されるスイッチアーキテクチャ400の高レベル構成について記載する。

#### 【0019】a 通知メモリー書き込み

PCIインターフェース404内のプリフェッチ可能なアドレス空間に対するPCIメモリー書き込み(MW)アクセス要求の受信に応答し、TWQ406内のメモリー空間が使用可能であれば、PCIトランザクション情報はTWQ406内にロードされる。TWQ406内に使用可能な空間がない場合には、TWQ406内にうまく配置されるまでそのトランザクションは再試行される。次に、TWQ406内部のロジックがトランザクションバケットを構築し、そのバケットを指定されたトランザクションキューQ1に送信する。キューQ1は、図3に図示されるように、ターゲットバスBの通知メモリー書き込み(PMW)キューとして指定されているものである。当該バケットは、こうして、トランザクションバケットのトランザクションヘッダーに記述されている通り、ターゲットバスBに経路指定される。

【0020】PCIインターフェース404若しくは418に入ってくるバケットは、各々そのメッセージヘッダー内にシーケンス番号を有している。シーケンス番号は、PCI書き込みトランザクションをそれが生成するPMWバケットに関連付ける。第2のシーケンス番号が各々のソースバス用に保持される。この追加のシーケンス番号は、ソースPCIバスインターフェースで書き込みトランザクションが実行される度に増分される。ソースバスにおける単一のPCI書き込み操作に関連付けられているシーケンスの中の最後のPMWは、1に設定された、最終PWMシーケンス指標も有する。ソースバスの読み出しトランザクションであって、同じバス上のPMWトランザクションと同じ方向に伝送されている、読み出しトランザクションは、TWQ406内部のメッセー

ジヘッダー内から上記シーケンス番号を取り出す。ターゲットバスBでの到着後、読み出しトランザクションは、それ自身のシーケンス番号と同じか、より小さいシーケンス番号を有する全てのPMWトランザクションを待ち、その後読み出しトランザクションの処理が許可される。このようにして、非シリアル転送において、PCIインターフェースに対するデータ順序付けが維持される。

【0021】ソースバスAにおいて転送が完了し、当該トランザクションに関連するデータバッファ（バッファ408）内の全てのデータがパケット化されてスイッチ装置に送信されたら、そのトランザクション情報はTWQ406から削除される。

b 遅延読み出し要求及び遅延読み出し完了  
スイッチアーキテクチャ400は、PCI I/O読み出し要求、構成読み出し、メモリー読み出し（MR）、MRL若しくはMRMに応答する。読み出し要求のための入力パケットとともにエンコードされたトランザクション情報は、現在TWQ406内部で処理されているトランザクションのトランザクション情報と比較される。一致した場合、スイッチアーキテクチャ400内部に、その入力パケットに関連したトランザクション用のアクティブな遅延読み出し要求（DRR）があることになる。

【0022】DRRアクティブ（一致検出）の場合、データ使用可能フラグがチェックされ、当該トランザクション用のデータが現在データバッファ408に存在するかどうか決定される。そのようなトランザクションデータがデータバッファ408内にある場合には、そのデータは、要求元マスターに転送される。そのようなトランザクションデータがなデータバッファ408内に無い場合には、トランザクションは再試行される。入力パケット（トランザクション要求）内にエンコードされたトランザクションデータがTWQ406内部で現在処理されているどのトランザクションとも一致しない場合には、スイッチアーキテクチャ400の応答は、TWQ406内のキューの現在の使用可能状況に依存する。作業キューが使用可能ではなくかつデータバッファ408内部に使用可能なバッファ空間が無い場合（スイッチビジー状態）には、入力されるトランザクション要求（読み出し要求）は後に再試行される。TWQ406内部でワークキューが使用可能であるか、データバッファ408内部でバッファ空間が使用可能である場合には、PCIトランザクション要求は、当該トランザクションに係る制御情報（コマンドアドレス、バイトインネブル）がTWQ406内に保存されて、当該トランザクションは再試行される。

【0023】TWQ406にトランザクション制御情報が追加された場合、スイッチアーキテクチャ400はトランザクションパケットを構成し、そのパケットをデ

スティネーションバスBの遅延トランザクションキュー（遅延読み出し要求Q2）に転送する。PMWシーケンスからのシーケンス番号が、このキューイングされたトランザクションパケットに割り当てられる。そしてこのトランザクションパケットは、プログラム可能なフローキュー302内部のロジックによって、トランザクションパケットヘッダーに記述されている通りターゲットバスBに経路指定される。DRRに対するメッセージペイロードは空である。メッセージヘッダーのコピーはTWQ406内に保持され、データバッファ408内部にメモリー空間が割り振られる。これらリソースは双方とも、関連するDRCがターゲットバスBから受信された場合に利用される。以下に列挙される状態のうちの一つが検出されるまで、追加のPCI読み出しコマンドがサービスされ得る。（1）TWQ406及びデータバッファ408の一方若しくは双方内の空間の全てが既に割り振られた。（2）ターゲットバスB用Q2（DRR）キューが満杯である（スイッチコアの受信許可により通知される）。

【0024】DRRがQ2の入力に到達すると、プログラム可能フローキュー302内のシーケンサーロジックは、当該DRRのシーケンス番号とターゲットに既に到着しているPMWにより運ばれてきたシーケンス番号及び「シーケンス内の最後のPMW」フラグとを比較する。全てのPMWが要求された通り終了し、全ての書込みデータがターゲットに送信されている場合、シーケンサーロジックは、ターゲットバスBにおいて要求を開始し、メモリーから1キャッシュラインを読み込む。当該データをターゲット（この場合はメモリー）から受け取ったら、DRCパケットが構成され、ソースバスAに関連付けられているQ3（DRC/DWCキュー）に転送される。PCIコマンドがMRMの場合、プログラム可能フローキュー内のシーケンサーロジックは、PCIマスターに代わって、メモリーから追加のキャッシュラインを読み出すための要求をバス上で開始することによって、データをプリフェッチする。ターゲットバスシーケンサーは、マスターがPCI転送をアクティブに保持する限りデータプリフェッチを続ける。Q3内の開始バスにおいてDRCパケットが受信されると、当該パケットがキューの先頭に到達した時に、そのパケットは除去される。メッセージヘッダー内の情報は、割り当てられたTWQ406にマッチされ、メッセージペイロード中のデータは、既に割り振られているデータバッファ408内に移される。

【0025】ターゲットバスBインターフェース418に関連付けられているロジックは、ソースバスA内のデータバッファが適当な空間を有さない限りデータをソースバスAに送信できない。ターゲットバスBインターフェース418内のプリフェッチロジックは、ソースバスAインターフェース404内のバッファに追加のデ

ータを送ることのできる時間について、ソースバスAロジック404と折衝する。優先度キューQ4は、高信頼性の媒体として用いられ、Q4を介して折衝メッセージが転送される。図6は、本発明の好適な実施例による、プログラム可能パケット出力キューイングのブロック図である。図6に示されるように、スイッチングインターフェース500は、プログラム可能フローキュー504を含んでおり、キュー504は、スケジューラ502と通信している。プログラム可能フローキューは、例えば図3及び4に定義されるような、複数のパラレル論理フローキューを有する。図6に更に詳しく示される通り、キューしきい値入力506及びデータ順序付け入力508はスケジューラ502に設定される。スケジューラ内のプログラム可能ロジックは、キューしきい値入力506及びデータ順序付け入力508からの入力を、プログラム可能フローキュー504に対するスケジューリングコマンドに翻訳し、その結果、スイッチングインターフェース500から出力されるパケットがプロトコル要件及び実時間キューステータスに従って転送される。

【0026】本発明の好適な実施例は、ここに記載された種々の方法を実行するためにプログラムされたコンピュータシステムとして、及びプログラム製品としての実施例を含むものである。コンピュータシステムにより本発明を実施する場合、開示された方法及びシステムを実行するための命令群は単一若しくは複数のコンピュータシステムのROM若しくはRAMのようなストレージ装

置内に配置されることになる。コンピュータシステムから要求を受けるまで、当該命令群は、他のコンピュータ記憶装置、例えば、ディスクドライブ（これにはディスクドライブに必要な応じて挿入して利用される、着脱可能な光ディスクやフロッピー（R）ディスクが含まれる）内にコンピュータプログラム製品として格納されることになろう。以上、本発明を特定の形で示し、好適な実施例について説明をしてきたが、当業者であれば、本発明の概念及びその範囲を逸脱することなく、形式及び詳細に関し多様な変形を成し得ることは容易に理解できるであろう。

#### 【図面の簡単な説明】

【図1】本発明によるフローアーキテクチャが実装された通信スイッチを示す図。

【図2】本発明の好適な実施例によるスイッチ内パケット構造を示す図。

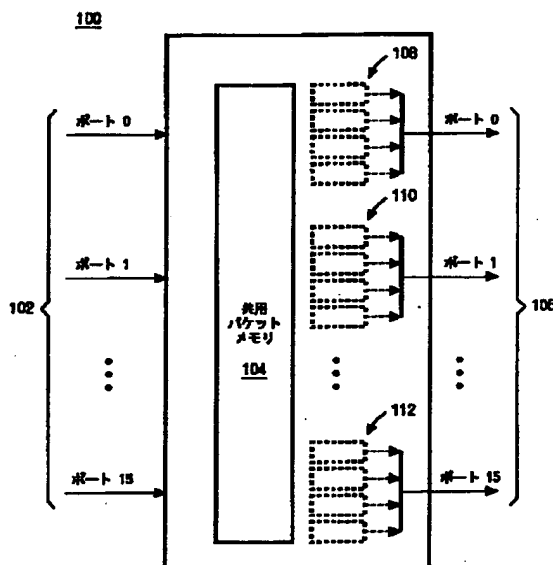
【図3】本発明の好適な実施例による、物理ポートに対してパケットをキューイングするための論理的ポートアダプターを示す図。

【図4】図3に示される論理ポートアダプター内部に実装することができるバスアーキテクチャ規則を示す説明図。

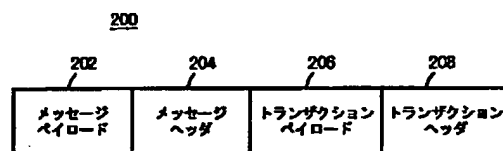
【図5】本発明の好適な実施例によるスイッチアーキテクチャのブロック図。

【図6】本発明の好適な実施例によるプログラム可能なパケット出力キューイングを示すブロック図。

【図1】



【図2】

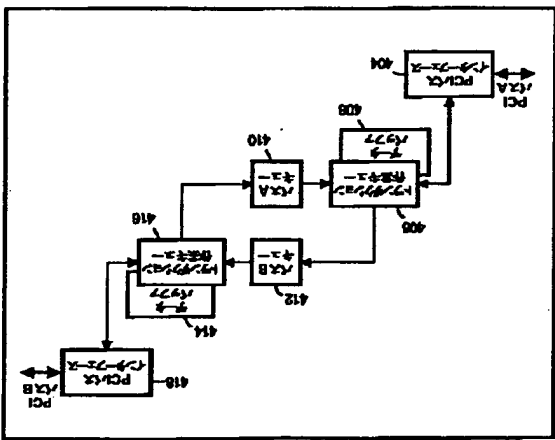


【図4】

行 → 列 ?	PMW	DRR	DWR	DRC	DWC
1. PMW	NO	YES	YES	YES	YES
2. DRR	NO	YES	YES	YES	YES
3. DWR	NO	YES	YES	YES	YES
4. DRC	NO	YES	YES	YES	YES
5. DWC	YES	YES	YES	YES	YES

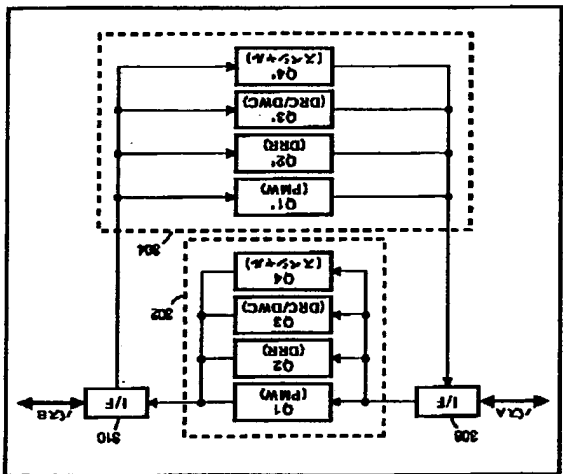


【図5】



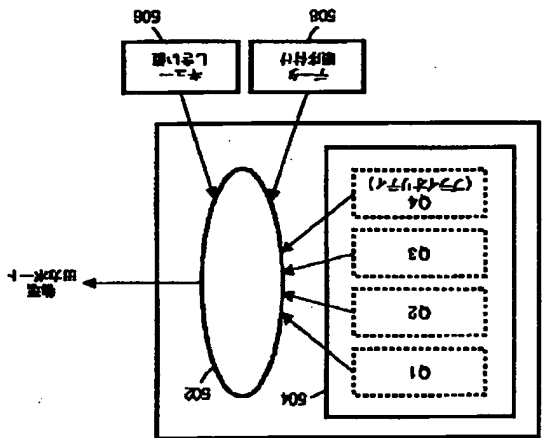
400

【図3】



300

【図6】



500

フロントページの続き

(72)発明者 メテイン・アイドミア  
アメリカ合衆国27713、ノースカロライナ  
州、ダーハム、サバルサック・コート 24  
州、ジョニー・ジョーン・リンス、サード  
アメリカ合衆国27516、ノースカロライナ  
州、チャペル・ヒル、カルデロン・ストリ  
ート 204

(72)発明者 ジョルジュ・ロラン・ロドリゲス  
アメリカ合衆国27511、ノースカロライナ  
州、ケプリー、エバート・ドライブ 404  
Fターム(参考) 5B061 FF00  
5K030 GA01 HA08 HB13 HB14 KA05  
KA13 KX02 KX11 KX18 KX24  
LC01